ComputerWeekly.com

Wednesday 31 May 2006

Technology

Servers

Itanium 2 Montecito to be multithreaded

Monday 17 November 2003

Intel is to introduce multithreading and 24Mbyte on-chip caches to its Itanium 2 processor family.

Itanium's first multicore processor, codenamed Montecito, will also have multithreading when it is released in 2005, said Ajay Malhotra, the director of enterprise marketing and planning for Intel's Enterprise Systems...

... group. The chip will also have 24Mbytes of on-chip cache, much more than the 9Mbytes available in Intel's Madison processors.

Multithreading allows a single processor to operate like multiple processors, so with multithreading, a dual-core Montecito would appear to the operating system to have at least four processors.

Intel's follow-up to Montecito, the 8-core Tanglewood processor being developed by a team of former Alpha processor developers, will also be multithreaded, but it may take a different approach to the problem. Malhotra said it will have seven times the performance of Madison Itanium 2 processors, an improvement that will primarily be driven by how Intel implements multithreading and how the company manages to link Tanglewood's cores together.

Multithreading is particularly difficult to bring to the Itanium "in order" processor architecture, said Nathan Brookwood, an analyst with industry research firm Insight 64.

"I'm amazed that they can do it because the secret for Itanium performance is very sophisticated compilers and multithreading is not the sort of stuff that compilers can accommodate."

Intel's move to 24Mbytes of on-chip cache is less remarkable, he added. "They're taking two Madison 9Mbytes and smashing them together on a chip, so you knew it was going to be at least 18Mbytes."

Robert McMillan *writes for IDG News Service*

ns, advises Gartner

Related Articles

- **EU court of justice blocks data transfer to US**

DEFENDANT'S EXHIBIT A

Insight > Hardware Friday 26th May 2006

Intel: Montecito or bust?

Rupert Goodwins in San Francisco
ZDNet UK
September 09, 2004, 10:40 BST

TALK BACK!
Tell us your opinion

IDF: The general manager of Intel's Enterprise Platform Group talks about the performance of its latest dual-core Montecito chip

Following Intel's first public demonstration of its latest Itanium -- the dual-core Montecito -- Abhi Talwalkar, general manager of Intel's Enterprise Platform Group, talked to ZDNet UK about future plans for the technology, how Itanium is currently being targeted and how the market for the chip will develop.

If you drop in a Montecito to a Madison-based system you claim around a 2x performance increase. Can we expect that sort of performance with each new generation of Itanium?
That's without recompilation -- you can get better than that if you recompile. For future performance increases, you have to look at Itanium history. From Merced to McKinley was a two times increase - that was an architectural change. From McKinley to Madison was 1.5x, a process shrink, but Madison to Montecito is an architectural change again with dual core and multi-threading and we're back to 2x.

It all depends on whether there's an architectural change or a process shrink. Montecito was a ground-up redesign to make it the best multi-core architecture. The cache size

DEFENDANT'S EXHIBIT B
Blumberg No. 5114

increase is another architectural change, but you'll see other changes beside that. There are good questions about what do you do if you have more than two cores -- what's the cache structure then? Montecito's cache is 24 megabytes (MB), it's really two independent caches of 12MBeach, fed by one bus interface. It's a pretty big bus.

How will Intel's lack of a chipset for larger servers affect the market?
We have our own chipset for two-way and four-way, but a lot of the development is greater than four-way and there the OEMs are developing their own chipsets. Most of these guys are RISC vendors, and they've been making investment in the RISC sides, and that's been more and more shipping over to Itanium2. NEC's a great example. NEC made a decision six or seven years ago to decrease investment on their proprietary microprocessors large scale systems and increase commensurately on It2. There'll be no white box activity above four-way Itanium2.

Won't that space be filled by Xeons?
There are 16 and, I think, 32 way Xeons in that space. Most of the scale-up work is in the Itanium2, because of our architectural advantages in the highly parallel EPIC architecture. We scale up a lot better, we have virtualisation that'll show up first in Montecito, a great technology for scaling up platforms. We have RAS advantages [reliability, availability and serviceability] in the silicon itself.

Next page

- Intel's Itanium: A leap too far?
- Intel prepares dual-core demo
- Dual-core chips will not share cache
- Microsoft puts back first 64-bit Windows
- Intel puts new server chips on roadmap
- AMD and Intel gear up for dual-core demos
- Microsoft wants mainframe to run Windows
- Intel reveals Itanium details

Email this Print this

TALK BACK

TALK BACK!
Tell us your opinion

Post your message here

Also in Insight
- Taking on the Office online
- RIAA's next moves i Washington
- 10 questions to ask technology solution i deployed by non-IT
- My friend, the robot
- Symantec: Beware t Microsoft 'monocultu
- Why you should car WiMax
- SAS's Goodnight wc goodbye yet
- Berners-Lee looks fc big leap
- The Windows Vista
- Keeping Microsoft li kicking

More...

REAL WORLD TECHNOLOGIES — In-Depth Technical Analyses And Help For Professionals

Sizing up the Super Heavyweights

By: Paul DeMone

Updated: 10-04-2004

Print | Email Author

Pages: Montecito - Make it a Double and Supersize It Go

Montecito – Make it a Double and Supersize It

At the beginning the 1990s Intel broke the million-transistor threshold with two 32-bit MPUs, the 486 and the 860. When Montecito ships next year, it will have taken Intel only 15 years to have crossed three orders of magnitude and cruise by the billion-transistor mark. The upcoming dual core 90nm IPF MPU packs 1.72 billion transistors on a single monster die largely because of its 26.5MB of integrated L2 and L3 cache. Based on analysis of public relations photos of 300mm Montecito wafers and recently disclosed die microphotographs it appears that Montecito is roughly 20 x 29mm or 580mm2 in size. The relative die size and floorplan of the Montecito is shown in Figure 3 along with those of its 130nm single core IPF predecessors Madison 6M and Madison 9M as well as fellow 90nm Intel chips Prescott and Dothan.

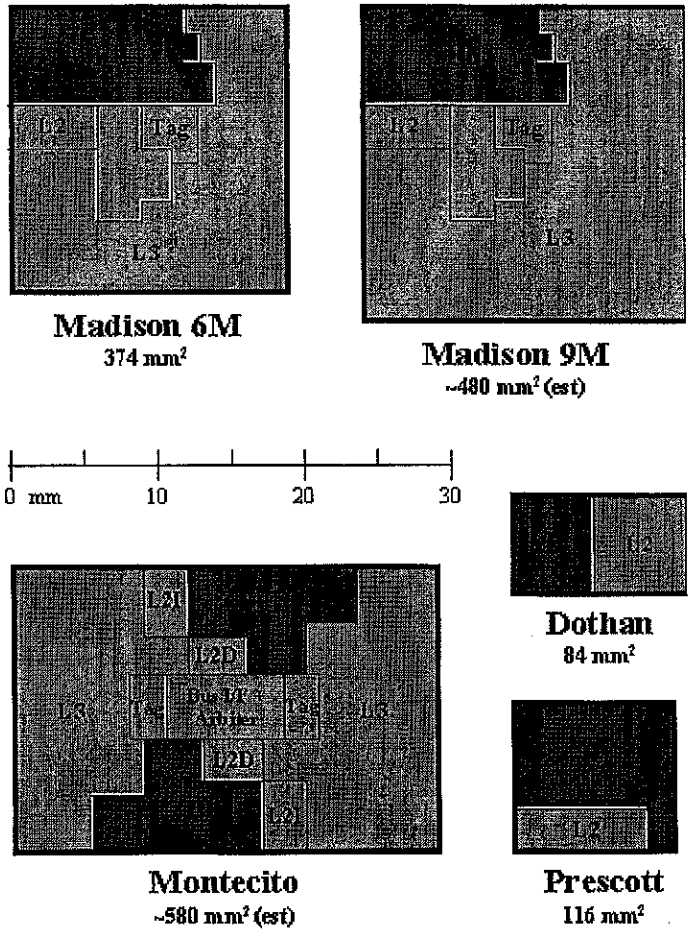

Figure 3 – Floorplan and Relative Die Size of Montecito and other Intel MPUs

Despite its size, Montecito will likely cost about the same or less than the estimated ~$125 of the existing Madison 6M to manufacture. Silicon cost is based on wafer count and varies weakly with wafer size. Montecito is manufactured on 300mm wafers with nearly 100 candidate dice per wafer, more than the ~63 Madison 6M on a 200mm wafers. Yield is not a specific issue for large memory intensive chips like Montecito, because it is over 70% L2 and L3 cache by area, and regular memory structures are protected against the majority of random point defects using redundant circuit and array elements. The Montecito CPU cores are only about 60mm2 each. The two CPU cores along with about 60mm2 of shared logic total about 180mm2 of non-cache region vulnerable to any defects, about the same critical area as a Willamette based P4 or Celeron.

The Montecito is more than simply dual Itanium 2 CPUs with more cache. Each CPU also incorporates coarse grained multithreading (CMT) in which hardware provides architected processor state for two threads along with logic to automatically switch execution from one thread to the other when the a thread relinquishes the CPU under software control or experiences a high latency event, like an L3 miss. The thread switch time is reportedly 15 cycles, which suggests a full pipeline flush is performed when switching threads. Although this sounds like a significant latency, one must keep in mind that for a 2+ GHz processor like Montecito an L3 miss could otherwise stall a CPU for 20 times longer or more. In addition to CMT, each Montecito CPU implements new IA-64 instructions, has extra functional units for shifting and population count, more efficient speculation recovery, and features for processor virtualization and enhanced reliability, availability, and serviceability (RAS) [5].

Although the cache hierarchy of previous IPF MPUs are arguable the most advanced of any processor family in terms of latency, bandwidth, and capacity, this was nevertheless an area of major improvement in Montecito. The changes are shown in Figure 4.

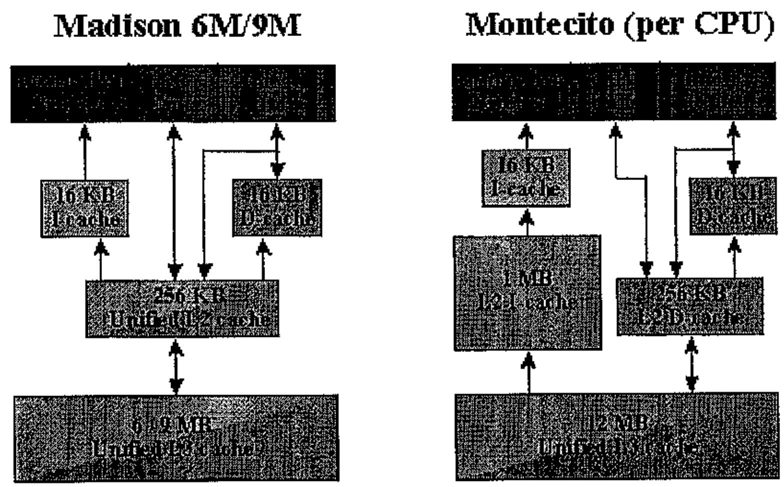

Figure 4 – Improvements in Montecito Cache Hierarchy

The biggest change was to split the unified 256KB L2 cache of the McKinley and Madison into separate 1MB L2 instruction cache and 256KB L2 data cache. This was done basically to eliminate instruction stream competition for the bandwidth and capacity of the L2 data cache. The 16KB instruction caches of the Madison and Montecito hold only 1024 instruction bundles which represents about 2.4k useful instructions taking into account the ~20% structural NOP content of a typical IA64 executable. To put that into perspective, that is only about 1/7th the instruction capacity of the POWER5's 64KB instruction cache. Obviously for many classes of programs, instruction stream fetching will represent a significant portion of the processor requests on the unified 256KB L2 as well a large portion of its contents.

By splitting the L2 caches in Montecito a lot of good things happen. From the data stream perspective, the 256KB L2 suddenly has one less port, and its entire 256KB capacity is available for data. This means less contention and stalls and fewer capacity and conflict misses. This adds up to more predictable memory hierarchy behavior, a very important feature for an architecture that relies heavily on static instruction scheduling. From the instruction stream perspective, the L2 I-cache can be located physically close to the L1 I-cache and its design optimized for the task. It doesn't need to be multi-ported or support sub-word access. As a result the 1MB L2 I-cache in Montecito likely has little or no latency penalty over the 256KB L2 D-cache, despite having four times its capacity. The combination of a very fast latency (1 cycle)

L1 I-cache and large and fast L2 I-cache has operational characteristics are impossible to duplicate in a single level cache. For example, if 90 % of instruction stream accesses that hit in the L1 or L2 hit in the L1, and the L2 has a latency of 6 cycles, than the L1/L2 combination performs like as a single level 1MB instruction cache with average latency of 0.9*1 + 0.1*6 = 1.5 cycles. This is half the latency of the 64KB instruction cache in the Alpha EV6/7 and AMD K7/8.

Information about the L3 caches in Montecito is encouraging but ambiguous. Although L3 capacity is doubled per CPU (12 MB) and quadrupled per device (24 MB) compared to Madison 6M, and L3 latency was said be the same as in Madison 6M and 9M [5]. The ambiguous part is that the Madison 6M's latency can be described either in absolute terms (9.3 ns) or in processor clock cycles (14). Given the size of the Montecito and the fact that it uses only seven layers of interconnect, just maintaining 9.3ns latency while doubling cache size is a good accomplishment. Keeping L3 latency at 14 cycles while clocking 50+% faster than Madison 6M (i.e. ~6 ns) would be an astounding accomplishment. Rounding out the improvements in Montecito's cache hierarchy are more efficient L2 and L3 queuing logic as well an increase in the number of L2 and L3 cache line victim buffers.

<< Prev Page Next Page >>

Pages: Montecito - Make it a Double and Supersize It Go

- United States
- Worldwide

- About Intel
- Press Room
- Contact Us

Search

- Products
- Technology & Research
- Solutions & Services
- Resource Centers
- Support & Downloads

- Press Room
- Browse News Resources
- News Releases
- by Date
- by Category
- Media Kits
- Photos
- Video / Audio
- Broadcast Content
- Media Contacts
- Find Corporate Information
- Corporate Information
- Biographies

Intel News Release

Forthcoming Dual-Core Intel® Itanium® Processor Achieves Fastest Four-Way Floating Point Benchmark

SANTA CLARA, Calif, July 7, 2005 – Based on internal testing by Intel Corporation, a system based on the forthcoming dual-core Intel® Itanium® processor codenamed "Montecito" demonstrated a 60 percent performance increase over a previous technical computing benchmark posted by a four-way RISC-based system.[1]

Using the LINPACK benchmark, which measures floating point performance, a system with four dual-core Itanium processors exceeded 45 GFLOPs (gigaflops), a measure of computer speed where a gigaflop is 1 billion floating-point operations per second. The previous record was 27.5 GFLOPs.[1]

"This performance result gives a peek into the advantage Montecito is expected to have over previous generations of the Itanium architecture for high-

Related Links

- More in this category
- Corporate Information
- Contact Corporate Press Relations

performance computing applications," said Phil Brace, general manager of Intel's Server Platform Group. "Three years ago we showed a four-processor Itanium-based system at 11.43 GFLOPs, and two years ago we hit 22.7 GFLOPs.[2] We are approaching the ability to reach a TeraFlop in as few as a 20-server system cluster and helping to dramatically increase the affordability to the scientific community."

Platforms using Montecito are expected to deliver up to twice the performance, up to three times the system bandwidth, and over 2 1/2 times as much on-die cache as the current generation of Itanium processors. While boosting performance, Montecito is expected to also deliver more than 20 percent lower power than previous generations of Itanium processors through new technologies for power management. Montecito will also have Intel® Hyper-Threading technology, enabling four times the threads as the current generation.

Intel, the world leader in silicon innovation, develops technologies, products and initiatives to continually advance how people work and live. Additional information about Intel is available at www.intel.com/pressroom.

1 Source: Intel Corporation, June 20, 2005. System Configuration: Intel Server Platform SR870BN4 using four Montecito Itanium 2 processors. Source: IBM Corporation, June 21, 2005: Best LINPACK based RISC result posted to http://www-1.ibm.com/servers/eserver/pseries/hardware/system_perf.pdf.

2 Source: HP, July, 8 2002: HP server rx5670* result of 11.43 GFLOPS using four Itanium 2 processors at 1.0 GHz with 3 MB L3 cache, HP-UX* 11i. Itanium® 2 processor 6M result of 22.7GFLOPs measured by Intel on Intel Server Platform SR870BN4 using four Itanium® 2 processors 6 MB L3 cache at 1.5 GHz.

Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance. Buyers should consult other sources of information to evaluate the performance of systems or components they are considering purchasing. For more information on performance tests and on the performance of Intel products, reference www.intel.com/performance.

- United States
- Worldwide

- About Intel
- Press Room
- Contact Us

- Search

- Products
- Technology & Research
- Solutions & Services
- Resource Centers
- Support & Downloads

- Press Room
- Browse News Resources
- News Releases
- by Date
- by Category
- Media Kits
- Photos
- Video / Audio
- Broadcast Content
- Media Contacts
- Find Corporate Information
- Corporate Information
- Biographies

Intel News Release

Intel® Itanium® 2 Processors Get Faster Bus Architecture

SANTA CLARA, Calif, July 18, 2005 - Intel Corporation today introduced two Intel® Itanium® 2 processors which deliver better performance over the current generation for database, business intelligence, enterprise resource planning and technical computing applications.

For the first time, Itanium 2 processors have a 667 megahertz (MHz) front side bus (FSB), which connects and transfers data between the microprocessor, chipset and system's main memory. Servers designed to utilize the new bus are expected to deliver more than 65 percent greater system bandwidth over servers designed with current Itanium 2 processors with a 400 MHz FSB. This new capability will help set the stage for the forthcoming dual core Itanium processor, codenamed "Montecito," which will feature the same bus architecture.

Intel Itanium 2 Processor

- Product Details: Intel® Itanium® 2 Processor
- Intel® Itanium® 2 Processor Family
- Benchmarks
- Developer
- Enterprise-Class Chipsets for Servers & Workstations

- Contact PR for Intel® Itanium® Processors

- Contact PR for Intel in the Enterprise

Blumberg No. 5114 DEFENDANT'S EXHIBIT E

"Intel continues to bring new capabilities to the Itanium architecture, evolving the platform to further improve performance for data intensive tasks," said Kirk Skaugen, general manager of Intel's Server Platforms Group. "Looking forward, we are coming up on the next major milestone for the Itanium processor family - dual-core server platforms based on Montecito. We are extremely excited about the customer and industry feedback we are getting, and the preliminary performance results we are seeing on Montecito-based systems will further expand Itanium's leadership in its targeted market segments."

Itanium-based servers continue to make strides in three target market segments: RISC replacement, mainframe migration and high-performance computing. Today, more than 40 percent of the Global 100 corporations have deployed Itanium-based servers and 79 of the TOP500 list of the world's fastest super computers are powered by Itanium processors. The ecosystem continues to grow with more than 3,600 applications available, while eight of nine RISC vendors and six of seven mainframe vendors sell mainframe-class Itanium-based servers.

The improved front side bus bandwidth allows for 10.6 gigabits of data per second to pass from the processor to other system components. In contrast, the current generation 400 MHz FSB transfers 6.4 gigabits of data per second. The ability to move more data in a very short period of time is critical to compute intensive applications in the scientific, oil and gas and government industries.

Hitachi, which will adopt the new Itanium 2 processors with the 667 FSB into new Hitachi BladeSymphony* servers coming in the next 30 days, has also designed a chipset (the communications controller between the processor and the rest of the computer system) to take advantage of the new bus architecture.

Platforms using Montecito are expected to deliver up to twice the performance, up to three times the system bandwidth, and more than 2 1/2 times as much on-die cache as the current generation of Itanium processors. While boosting performance, Montecito is expected to also deliver more than 20 percent lower power than previous generations of Itanium processors due to new technologies for power management. Montecito will also have Intel® Hyper-Threading technology, enabling four times the threads as the current generation.

The Intel Itanium 2 processor at 1.66 GHz with 9 MB of cache with 667 FSB is available for $4,655 in 1,000-unit quantities. The Intel Itanium 2 processor at 1.66 GHz with 6 MB with 667 FSB of cache will be available for $2194 in 1,000-unit quantities.

Intel, the world leader in silicon innovation, develops technologies, products and initiatives to continually advance how people work and live. Additional information about Intel is available at www.intel.com/pressroom.

back to top

- Site Map
- RSS Feeds
- Jobs at Intel
- Investor Relations

- *Legal Information
- Privacy Policy
- ©2006 Intel Corporation

Home > News > **Multithreading Set for Processors**

Multithreading Set for Processors

By Jeffrey Burt

March 3, 2003

Intel Corp., IBM and Sun Microsystems Inc. are each looking to incorporate multithreading capabilities into multiple-core processors.

Intel last month confirmed it will put multithreading, which it calls hyperthreading, into a future version of its 64-bit Itanium line of chips. The Santa Clara, Calif., company will have dual-core processing in an Itanium chip code-named Montecito and due in 2005.

Intel officials would not say when the vendor will add hyperthreading to the Itanium, but a source familiar with Intel's plans said the company aims to include it in Montecito. Last week, IBM—which has had dual-core processing in its Power4 chip since 2001—and Sun said they plan to add multithreading capabilities to their dual-core chips over the next couple of years.

Most currently available chips can process only a single thread of instructions at once, and a lot of time is spent waiting for memory. With multithreading, if a process is waiting for memory for one thread, the chip can work on another.

"This allows a single processor to do more work," said Rob Enderle, an analyst at Giga Information Group Inc., in San Jose, Calif. "You're able to pack more processing power in a smaller space."

In this way, enterprises can increase the workload of each server, which is key if they are going to be successful in their efforts to pack more processing power into their server farms. It could also mean cost savings to users, who will be able to do more work with fewer servers, Enderle said.

IBM will put the multithreading feature—the ability of a chip to run multiple tasks, or threads, at the same time—on its Power5 server chip, due for release next year, said Mark Papermaster, director of microprocessor design for IBM's Systems Group.

DEFENDANT'S EXHIBIT
Blumberg No. 5114

In the Power5, each core will be able to handle two threads simultaneously, a capability that when combined with other performance enhancements in the chip, will boost user performance fourfold, said Papermaster, in Austin, Texas.

Processor Performance Hits the Fast TrackX

Chip makers plan to build multithreading on multiple cores
- **Intel Will release dual-core processing in Montecito chip in 2005**
- **IBM Will release multithreading capabilities in its dual-core Power5 next year**
- **Sun Will release dual-core UltraSPARC IV later this year, multithreading capabilities in Niagara chip in 2005**

He said samples of the Power5, which will run faster than 1.5GHz, are performing well in internal tests. The chip will be important for such jobs as high-transaction applications or heavy data mining tasks, he said. It will also be able to handle single-threaded applications.

At its annual analyst conference in San Francisco last week, Santa Clara-based Sun began rolling out an aggressive road map for its UltraSPARC processors, including its first dual-core chip, the UltraSPARC IV, set for release later this year. In 2005, Sun officials said, the company will roll out a chip—code-named Niagara—that will be built on a 0.09-micron manufacturing process and will feature multithreading capabilities. The goal of Niagara's simple design will be to enable eight cores to run four threads simultaneously, according to Sun.

Multithreading on processors will be a key component of Sun's Throughput Computing chip strategy, designed to enhance application performance by maximizing network computing throughput, or the total amount of work done, said the company. The result will be faster application speed and a smaller system footprint, which will mean cost savings, Sun officials said.

Print Email Reprints

XML **Add** eWEEK Technology News to your RSS newsreader or My Yahoo!

TALKBACK

Sign In To Talkback! | Register

VA: Dat
than Re
6.7.200
Oracle E
Tender (
Software
6.7.200
eWEEK I
Podcast,
6.7.200
IBM to E
Source I
Style for

WHAT Y
READIN
Baselin
• Repor
Security
Drives I
• Micros
Recover
Becomi
• Moppi
• Cavali
Caught
• Tradir

CIO In
• CIO Ir
Booth o
Catchin
• Case S
Searche
Data for
Version
• Case S
Searche
Data for
Version
• CIO Ir
Booth o
Catchin
• Manag
Liaisons

eWEEK
XML
delivere
desktop

COMPA
Spa
Implem
Not Ha

New En
Compar
Ziff Dav